



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

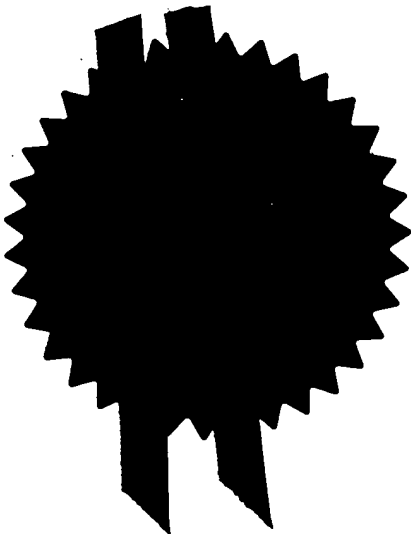
CERTIFIED COPY OF PRIORITY DOCUMENT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

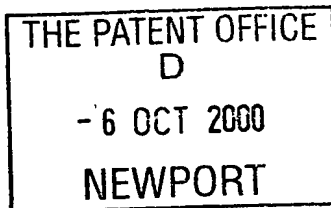
Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed:

Dated: 10 May 2001

THIS PAGE BLANK (USPTO)



060CT00 E573851-1 D00393
P01/7700 0.00-0024455.8

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference

2000P04862/GB/R76/MM/cs

2. Pate.
(The

0024455.8

6 OCT 2000

3. Full name, address and postcode of the or of each applicant (*underline all surnames*)

Roke Manor Research Limited
Old Salisbury Lane
Romsey
Hampshire
SO51 0ZN

Patents ADP number (*if you know it*)

If the applicant is a corporate body, give the country/state of its incorporation

Great Britain

561 545505

4. Title of the invention

IMPROVEMENTS IN OR RELATING TO PACKET SWITCHES

5. Name of your agent (*if you have one*)

"Address for service" in the United Kingdom to which all correspondence should be sent (*including the postcode*)

Siemens Shared Services Limited
Intellectual Property Department
Siemens House, Oldbury
Bracknell, Berkshire RG12 8FZ
United Kingdom

Patents ADP number (*if you know it*)

7761 000002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (*if you know it*) the or each application number

Country	Priority application number (<i>if you know it</i>)	Date of filing (<i>day / month / year</i>)
GB	0012600.3	25/05/2000

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing (<i>day / month / year</i>)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (*Answer 'Yes' if:*

YES

- a) any applicant named in part 3 is not an inventor, or
 - b) there is an inventor who is not named as an applicant, or
 - c) any named applicant is a corporate body.
- See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 10

Claim(s) 2

Abstract 1

Drawing(s) 5 ⁵

10. If you are also filing any of the following, state how many against each item.

Priority documents 0

Translations of priority documents 0

Statement of inventorship and right to grant of a patent (Patents Form 7/77) 3 ✓

Request for preliminary examination and search (Patents Form 9/77) 1 ✓

Request for substantive examination (Patents Form 10/77) 0

Any other documents 0
(please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature

Margaret Mackett
Margaret Mackett

Date 04/10/2000

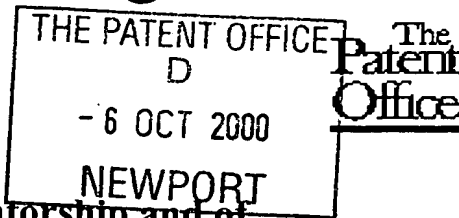
12. Name and daytime telephone number of person to contact in the United Kingdom
Margaret Mackett - 01344 396808

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- b) Write your answers in capital letters using black ink or you may type them.
- c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- e) Once you have filled in the form you must remember to sign and date it.
- f) For details of the fee and ways to pay please contact the Patent Office.



7/77

**Statement of inventorship and of
right to grant of a patent**

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

2000P04862/GB/R76/MM/cs

2. Patent no.

(If you, know it)

0024455.8

- 6 OCT 2000

3. Full name of the or of each applicant

Roke Manor Research Limited

4. Title of the invention

IMPROVEMENTS IN OR RELATING TO PACKET SWITCHES

5. State how the applicant(s) derived the right
from the inventor(s) to be granted a patent

By provisions of contract of service and Section 39(1) (a) and (b) of
the Patent Act 1977.

6. How many, if any, additional Patents Forms
7/77 are attached to this form?
(see note (c))

2

I/We believe that the person(s) named over the page (and on any
extra copies of this form) is/are the inventor(s) of the invention
which the above patent application relates to.

Signature

M Margaret Mackett
Margaret Mackett

Date 04/10/2000

8. Name and daytime telephone number of
person to contact in the United Kingdom

Margaret Mackett - 01344 396808

Notes

- a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- b) Write your answers in capital letters using black ink or you may type them.
- c) If there are more than three inventors, please give the names and addresses of the other inventors on the back of another Patents Form 7/77 and attach it to this form.
- d) When an application does not declare any priority, or declares priority from an earlier UK application, you must provide enough copies of this form so that the Patent Office can send one to each inventor who is not an applicant.
- e) Once you have filled in the form you must remember to sign and date it.

REEVE, Andrew
47 Western Road
Winchester
Hants
SO22 5AH

7632888001

Patents ADP number (if you know it):

DAVIS, Simon Paul
17 Westering
Romsey
Hants
SO51 7LX

5873773002

Patents ADP number (if you know it):

Patents ADP number (if you know it):

Have you signed the form?

IMPROVEMENTS IN OR RELATING TO PACKET SWITCHES

The present invention relates to improvements in or relating to packet switches, and is particularly concerned with the use of such switches in
5 Internet systems.

Traffic volume in the Internet is growing exponentially, almost doubling every 3 months. The current capacity of internet protocol (IP) routers is insufficient to meet this demand and hence there is a need for IP routers that can route IP traffic at extremely large aggregate bandwidths in
10 the order of several Terabit/s. Such routers are termed "Terabit Routers".

Two important trends are also evident. First, operators are consolidating all traffic onto a single IP back-bone. Secondly, IP is increasingly required to support real-time and multimedia traffic. This means that the next generation of routers must also support 'Quality of Service'
15 (QoS). In particular, they must support low bounded delay for real-time traffic.

It is therefore an object of the present invention to provide a traffic management system suitable for use in a terabit router which meets the requirements for IP traffic. In such a router, variable length packets of data
20 are sent in fixed length cells across a cross-bar switch.

In accordance with one aspect of the present invention, there is provided a traffic management system for a packet switch comprising:-

a cross-bar;

a plurality of ingress means connected to an input side of the cross-
25 bar;

a plurality of egress means connected to an output side of the cross-
bar;

a bandwidth controller for allocating a bandwidth to each ingress-egress pairing; and

a cross-bar controller for controlling operation of the cross-bar in accordance with the bandwidth allocated by the bandwidth controller.

5 A packet switch is a device which receives packets of data on a plurality of input ports and transfers each packet to a specific one of a plurality of output ports. The packets of data can be of variable length or fixed length. An example of a packet switch is a terabit router as described above.

10 In accordance with another aspect of the present invention, there is provided a method of controlling a packet switch connected between a plurality of ingress means and a plurality of egress means, each ingress means having a plurality of packet queues for transmission, the method comprising the steps of:-

15 a) defining a period over which the packet queues are to be transmitted;

 b) calculating a rate matrix having elements corresponding to the rates from an ingress means to an egress means;

 c) at the beginning of each period, calculating a cell matrix
20 containing a number of cells which must be transmitted from each of the packet queues during the period;

 d) for each cell slot in the period, determining a configuration which matches the cell matrix by only servicing packet queues with non-zero cell counts, the configuration being determined in accordance with the
25 following constraints:-

 (i) selecting no more than one cell from each ingress means;
 and

- (ii) routing no more than one cell to each egress means;
- e) decrementing the cell counts of each queue serviced by one; and
- f) repeating steps c), d) and e) until the end of the period.

The term 'servicing' is intended to mean selecting a queue,
5 transmitting a cell from that queue, and then deleting that cell from the queue.

The ingress means and the egress means may comprise line interface cards (LICs) which may also carry out a forwarding function. Such LICs may be known as 'forwarders'.

10 A traffic management system and its method of operation in accordance with the present invention advantageously is starvation free, and bandwidth is always allocated to a packet queue with non-zero occupancy across the cross-bar. The allocation of bandwidth and priority to any packet queue is also fair. Configured rates are also provided in accordance with
15 each ingress packet queue – the rates being calculated to ensure the QoS of traffic streams.

Moreover, the system and method in accordance with the present invention also provides bounded low delay across the cross-bar, and the efficient (substantially 100%) utilisation of the cross-bar without loss of
20 performance.

For a better understanding of the present invention, reference will now be made, by way of example only, to the accompanying drawings in which:-

Figure 1 illustrates a terabit router architecture;

25 Figure 2 illustrates a cross-bar controller for use in the Figure 1 router architecture;

Figure 3 illustrates a rate matrix (8x8);

Figure 4 illustrates an example cross-bar configuration matrix (8x8);
and

Figure 5 illustrates a 'find_config' procedure.

Terabit routers will require a scalable high capacity communications
5 path between its line functions. One technique of implementing this is to use
a cell based cross-bar as will be described below. However, the effectiveness
of this technique with respect to the efficiency of the router and its ability to
support the quality of service (QoS) of the communications it supports, will
depend on the cross-bar configuration method employed.

10 In accordance with the present invention such a cross-bar
configuration method provides very high levels of efficiency and support for
quality of service. It has application in any switch or router such as IP
routers, asynchronous transfer mode (ATM) switches or MPLS label switch
routers.

15 Although cross-bar architectures are already known for use in routers
and switches, they typically suffer from the inability to achieve 100%
utilisation of the cross-bar without incurring very long delays; and the
inability to guarantee upper bounds for delay as required to support real-time
traffic.

20 Initially, a terabit router architecture is described and an explanation
given of what is meant by a cell based cross-bar. Although the present
invention is described with reference to a terabit router, it will readily be
appreciated that it is not limited to such a device and is equally applicable to
any packet switch device as stated above.

25 Figure 1 illustrates a terabit router architecture 100 in which packets
arrive at ingress forwarders 102, 104, 106 via their input port(s) (not shown)
and are routed across a cross-bar 110 to a correct egress forwarder 120 which

transmits them across its output port(s) (not shown). Each ingress forwarder 102, 104, 106 maintains a separate packet queue for each egress forwarder 120.

It will be appreciated that although only one egress forwarder 120 is shown in Figure 1, the number of egress forwarders will be the same as the number of ingress forwarders.

By way of explanation, a cell based cross-bar is characterised as follows:

- a) Each ingress line function may be connected to any egress line functions.
- b) Each ingress line function may only be connected to one egress line function at a time.
- c) Each egress line function may only be connected to one ingress line function at a time.
- d) All ingresses transmit in parallel across the cross-bar.
- e) Data is transmitted across the cross-bar in small fixed sized cells, for example, a cell size is typically 64 octets.
- f) Cell transmission is synchronised across all the ingress line functions. This means that for each cell cycle, each ingress line function starts transmitting the next cell at the same time.
- g) The cross-bar is reconfigured at the end of every cell cycle.

The present invention relates to a cross-bar unicast cell scheduling method for which the following desirable features:-

- i starvation free: No packet queue with non-zero occupancy must ever be starved of bandwidth across the cross-bar.
- ii fair: The allocation of bandwidth and priority to packet queue must be fair.

- iii provide configured rates: Rates must be implemented which are configured for each ingress packet queue q_{jk} where j indicates the ingress, k indicates the egress, and q_{jk} represents the packet queue at the ingress j for packets destined for egress k . The rates are calculated to ensure the QoS of traffic streams.
- iv bounded low delay: Bounded low delay across the cross-bar must be provided. Note that the overall delay across the whole router will also be determined by the traffic management functions implemented in the line interface cards (LICs).
- v efficiency: It is desirable that 100% utilisation of the cross-bar be achieved without loss of performance.

An embodiment of a cross-bar controller arrangement 200 in accordance with the present invention is depicted in Figure 2.

The cross-bar controller arrangement 200 comprises a cross-bar 202 which is controlled by a cross-bar controller 204 which in turn is controlled by a bandwidth controller 206. The bandwidth controller 206 is responsible for efficient allocation of the bandwidth across the cross-bar 202, and calculates the rates that each ingress forwarder 210, 212, 214, 216 must transmit to each egress forwarder 220, 222, 224, 226. This is the same as the rate at which data must be transmitted from each packet queue. The means by which these rates are calculated is beyond the scope of this description.

The bandwidth controller 206 transmits the rates to the cross-bar controller 204 which is responsible for efficient scheduling of data across the cross-bar 202 whilst maintaining the rates calculated by the bandwidth controller 206. The cross-bar controller 204 is responsible for calculating the following information at the end of each cell cycle.

1. it transmits to each ingress forwarder the identity of the next packet queue from which to transmit. Recall that each ingress forwarder maintains a separate queue of packets for each egress forwarder; and
2. it transmits to the cross-bar the next cross-bar configuration.

5 A description of how the cross-bar controller 204 determines this information follows:-

The rates can be represented using a matrix as depicted in Figure 3, which provides the example of an 8×8 router. Let us call this matrix R , with elements r_{jk} , such that r_{jk} is the rate from ingress forwarder j to egress

10 forwarder k . The rate unit employed is cells per unit time.

Let F be the maximum rate at which an ingress forwarder 210, 212, 214, 216 can receive data from its input port and transmit it across the cross-bar 202. F is also the maximum rate at which an egress forwarder 220, 222, 224, 226 can receive data from the cross-bar port and transmit it across its

15 output port. This is the maximum possible rate independent of any traffic conditions prevailing at the time.

Let N be the number of ingress forwarders 210, 212, 214, 216. N is also the number of egress forwarders 220, 222, 224, 226.

20 Then, the following inequalities must be obeyed:-

For all j , where the row sums cannot exceed F :-

$$\sum_{k=1}^N r_{jk} \leq F \quad (1)$$

For all k , where the column sums cannot exceed F :-

$$\sum_{j=1}^N r_{jk} \leq F \quad (2)$$

25 In fact, the bandwidth controller 204 will ensure that the following hold true:-

For all j , the row sums are all equal to F :-

$$\sum_{k=1}^N r_{jk} = F \quad (3)$$

For all k , the column sums are all equal to F :-

$$\sum_{j=1}^N r_{jk} = F \quad (4)$$

5 and

$$\sum_{k=1}^N \sum_{j=1}^N r_{jk} = NF \quad (5)$$

Herein, a matrix with the property that the column and row sums are all equal shall be called a sum-perfect square.

The configuration of the cross-bar 202 can be depicted by a matrix such as that shown in Figure 4. Each entry in the matrix may take the value '0' or '1'. An entry with the value '1' indicates that the cross-bar 202 is configured for transmission of a cell from the corresponding ingress (row index) to the corresponding egress (column index). In particular, a value of '1' indicates a cross-connect between the ingress and egress, and a value of '0' indicates absence of a cross-connect.

In Figure 4, ingress 1 cross-connects with egress 1, ingress 2 with egress 2, ingress 3 with egress 4, ingress 4 with egress 8, ingress 5 with egress 5, ingress 6 with egress 6, ingress 7 with egress 3, and ingress 8 with egress 7.

20 For unicast operation, the cross-bar configuration matrix has the following properties:-

1. Each row contains exactly one non-zero entry. (Each ingress transmits a cell to exactly one egress.)
2. Each column contains exactly one non-zero entry. (Each egress
25 can receive a cell from one ingress only.)

Note that these properties mean that the matrix forms a sum-perfect square where the row/column sum is 1. Let us call this matrix X .

The steps of the method of operation are as follows.

- 1) The rate matrix R is re-calculated periodically.
- 5 2) The rate is maintained over the length of each period.
- 3) At the beginning of each period, calculate a matrix containing the number of cells that must be transmitted from each of the packet queues q_{jk} during the period to achieve its rate.
- 4) For each cell slot in the period, find a configuration which matches
10 the cell count matrix by only servicing queues with non-zero cell counts.
- 5) Decrement by 1 the cell counts of each queue serviced.
- 6) Repeat from 3) until the end of the period at which all cell counts will be zero.

15 This is expanded below:-

Let P be the length of the period. This is chosen to be a whole number of cell cycles and to be of magnitude no greater than the maximum tolerated delay for real-time traffic.

20 Let T be the cell slot number in the period P , such $T = 0$ is the first slot and $T = P - 1$ is the last.

Let C^T be the cell count matrix at the beginning of cell slot T . Then C^0 is calculated as follows:-

$$C^0 = PR \tag{6}$$

25 The rate matrix R is calculated such that C^0 contains integral values. One way of doing this is described in co-pending British patent application no. 9907313.2 (docket number F21558/98P4863).

Note that since R is sum-perfect, then so is C^0 .

For each T , calculate a cross-bar configuration matrix X^T such that

$$x_{jk}^T = 1 \Rightarrow c_{jk}^T > 0 \quad (7)$$

Since X^T provides a cell send opportunity for each q_{jk} such that $x_{jk}^T = 1$,

C^{T+1} is calculated as follows:-

$$5 \quad C^{T+1} = C^T - X^T \quad (8)$$

Since C^0 is sum-perfect and X^T is sum-perfect for all T , then it is easy to prove that C^T is sum-perfect for all T . In particular, C^P is sum-perfect with sum 0 so that all elements are zero and all queues have been given their full complement of cell send opportunities.

10 An example algorithm for finding a configuration matrix X^T matching cell count matrix C^T is presented below. The main routine calls the 'find_config' routine for the number of time slots in period P . 'find_config' is a recursive heuristic algorithm.

The main routine is:

15 **for** slot = 0 **to** period_P **do**
begin
 if not find_config(ingress = 0) **then abort;**
end

The 'find_config' routine is shown in Figure 5.

CLAIMS:

1. A traffic management system for a packet switch comprising:-
a cross-bar;
a plurality of ingress means connected to an input side of the cross-bar;
a plurality of egress means connected to an output side of the cross-bar;
a bandwidth controller for allocating a bandwidth to each ingress-egress pairing; and
a cross-bar controller for controlling operation of the cross-bar in accordance with the bandwidth allocated by the bandwidth controller.
2. A system according to claim 1, wherein the cross-bar controller also selects the next ingress-egress pairing for each ingress means.
3. A method of controlling a packet switch connected between a plurality of ingress means and a plurality of egress means, each ingress means having a packet queue for transmission, the method comprising the steps of:-
 - a) defining a period over which the packet queues are to be transmitted;
 - b) calculating a rate matrix having elements corresponding to the rates from an ingress means to an egress means;
 - c) at the beginning of each period, calculating a cell matrix containing a number of cells which must be transmitted from each of the packet queues during the period;

d) for each cell slot in the period, determining a configuration which matches the cell matrix by only servicing packet queues with non-zero cell counts, the configuration being determined in accordance with the following constraints:-

- (i) selecting no more than one cell from each ingress means;
and
- (ii) routing no more than one cell to each egress means;
- e) decrementing the cell counts of each queue serviced by one; and
- f) repeating steps c), d) and e) until the end of the period.

4. A traffic management system substantially as hereinbefore described with reference to Figure 2 of the accompanying drawings.

5. A method of controlling a packet switch substantially as hereinbefore described with reference to Figures 2 to 5 of the accompanying drawings.

ABSTRACT
IMPROVEMENTS IN OR RELATING TO PACKET SWITCHES

Described herein is a method of cell level scheduling for handling unicast traffic in routing devices, for example, crossbar switches. This is achieved by the provision of support for multimedia and real-time traffic in large bandwidth routing devices known as terabit routers. Each terabit router has a plurality of ingress line interface cards (210, 212, 214, 216), a plurality of egress line interface cards (220, 222, 224, 226) and a cell based cross-bar (202). Each ingress card has a plurality of queues, a different queue for each of the egress cards respectively. The cross-bar (202) is controlled by a cross-bar controller (204) in association with a bandwidth controller (206). In operation, a target rate matrix is maintained over a set period. At the beginning of each period, a matrix of numbers of cells queued to be transmitted is calculated in accordance with the target rate. For each successive cell slot within the period, a configuration is found which matches the cell number matrix by servicing only queues which have non-zero cell counts. At the end of the period, the cell counts are all zero.

(Fig. 2)

THIS PAGE BLANK (USPTO)

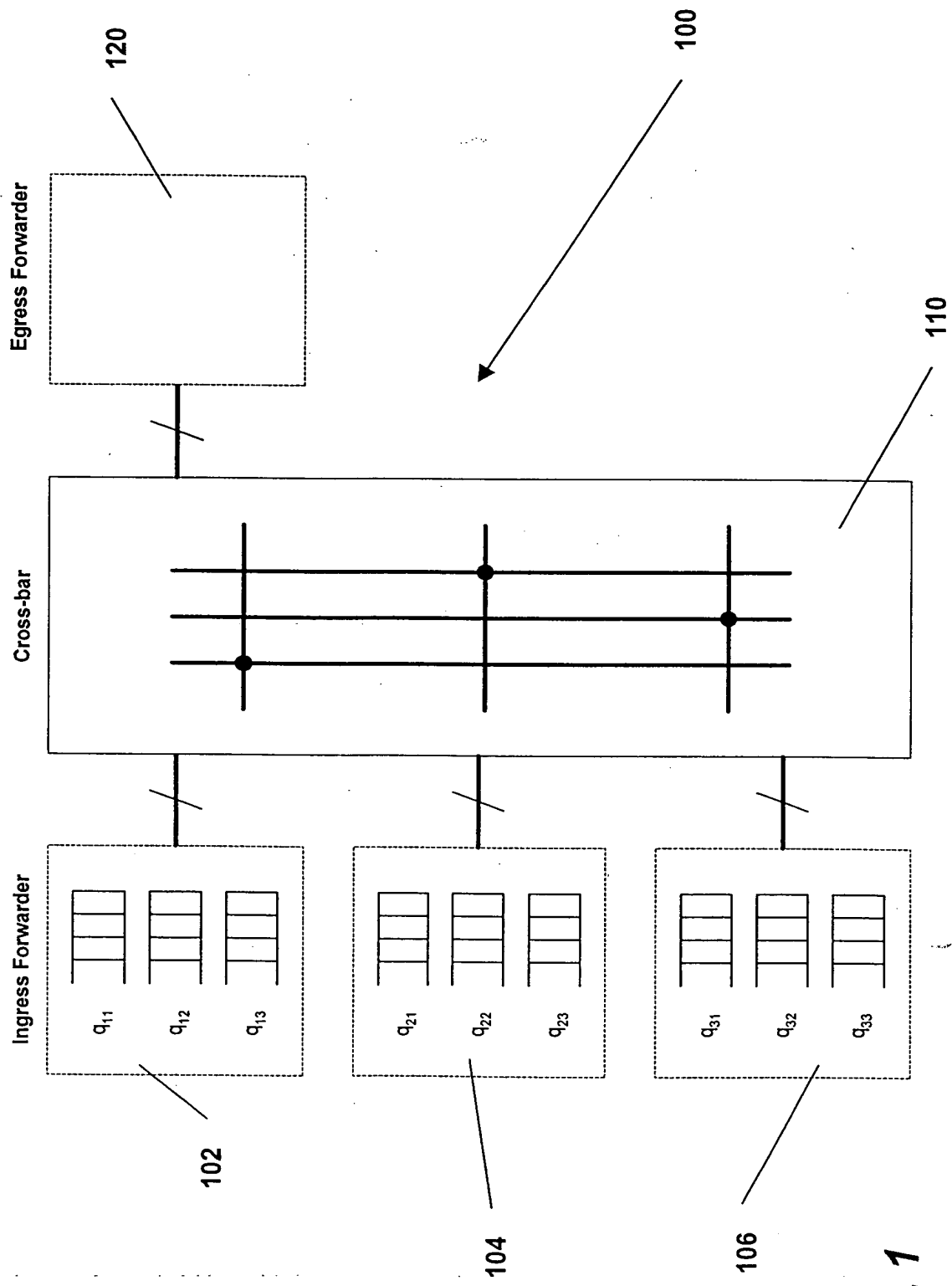


Fig. 1

THIS PAGE BLANK (USPTO)

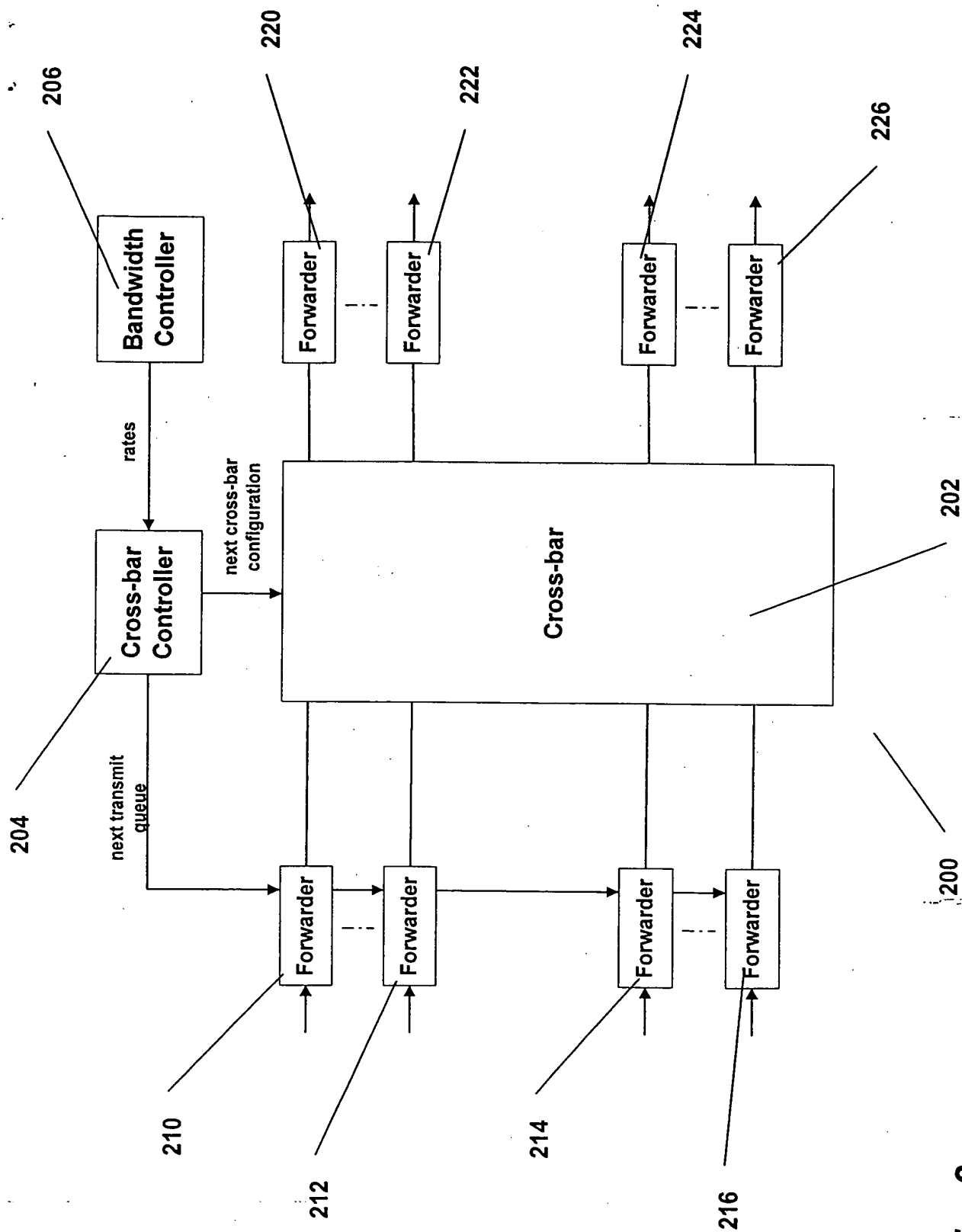


Fig. 2

THIS PAGE BLANK (USPTO)

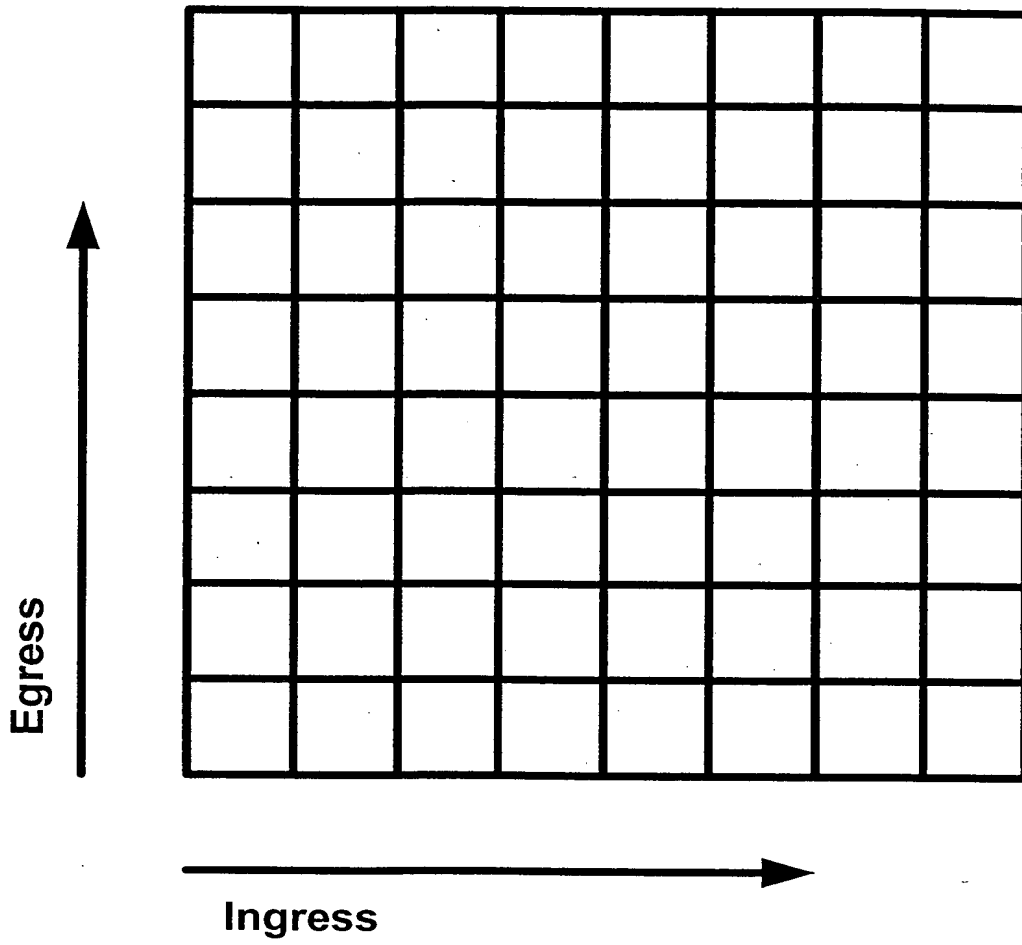


Fig. 3

THIS PAGE BLANK (USPTO)

Egress ↑

1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0

↑ Ingress

Fig. 4

THIS PAGE BLANK (USPTO)

```
find_config(ingress) returns boolean
begin
    for all egress ports do
        begin
            if ((cell_count[ingress, egress] > 0) and picked[egress] == false) then
                picked[egress] = true;
                if ((ingress == last ingress) or find_config(ingress+1)) then
                    return true;
                else
                    picked[egress] = false;
                end
            end
        end
    end
    return false;    // no match configuration found
end
```

Fig. 5

THIS PAGE BLANK (USPTO)